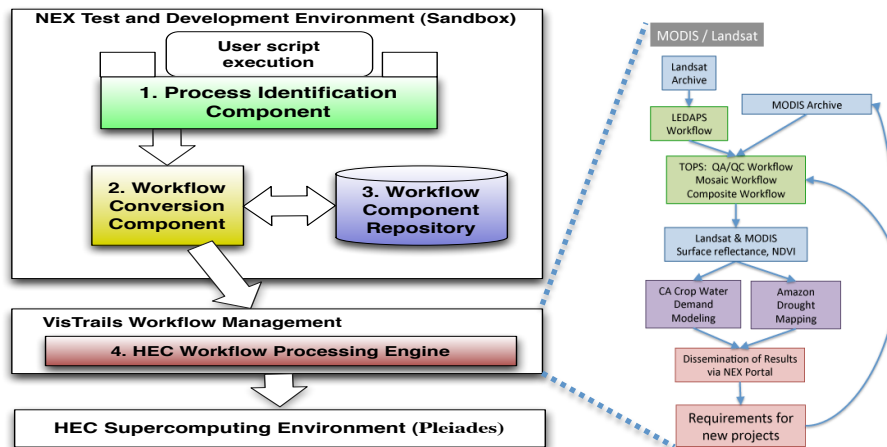# Semi-automatic Science Workflow Synthesis for High-end Computing on the NASA Earth Exchange (NEX)

PI: Ramakrishna Nemani, NASA ARC

Objective

Enable scientists to capture workflows (sequences of analysis tasks) and workflow provenance to significantly enhance reproducibility, transparency, and sharing of scientific results on NEX

- Develop components for automatic workflow capture, archiving and management.
- Enable workflow migration between local computer systems and the supercomputer environment in NEX.

Architecture implements user scripts converted to workflows managed by VisTrails running on supercomputing environment.

Approach

- Develop components for process capture from user scripts, automatic workflow generation and archiving.
- Integrate the workflow components with VisTrails data and process provenance management system.
- Develop workflow engine for VisTrails that will interface with the NASA supercomputing environment.
- Test the end-to-end system with workflows for data analysis and modeling.

Co-Is/Partners: Petr Votava, Andrew Michaelis, Hirofumi Hashimoto, Weile Wang, California State University Monterey Bay

Key Milestones

• Complete preliminary design	09/12
• Complete process identification component	05/13
• Develop initial Portable Batch System interface	07/13
• Complete workflow conversion component	01/14
• Complete workflow repository	03/14
• Complete initial supercomputer interface	07/14
• Finish component integration	01/15
• Complete science integration and testing	05/15

TRL_{in} = 3 $TRL_{current}$ = 4

Earth Science Technology Office